*Protein Science* (1992), *1*, 590-600. Cambridge University Press. Printed in the USA.


# Amino acid sequence of fibrolase, a direct-acting fibrinolytic enzyme from *Agkistrodon contortrix contortrix* venom

ANNE RANDOLPH,[1] SCOTT H. CHAMBERLAIN,[1] H.-L. CATHY CHU,[1]
ANASTASSIOS D. RETZIOS,[2] FRANCIS S. MARKLAND, JR.,[2]
AND FRANK R. MASIARZ[1]

[1] Chiron Research Laboratories, Chiron Corporation, 4560 Horton Street, Emeryville, California 94608
[2] Department of Biochemistry and Comprehensive Cancer Center, University of Southern California School of Medicine, 1303 North Mission Road, Los Angeles, California 90033

(RECEIVED October 2, 1991; REVISED MANUSCRIPT RECEIVED December 19, 1991)

**Abstract**

The complete amino acid sequence of fibrolase, a fibrinolytic enzyme from southern copperhead (*Agkistrodon contortrix contortrix*) venom, has been determined. This is the first report of the sequence of a direct-acting, nonhemorrhagic fibrinolytic enzyme found in snake venom. The majority of the sequence was established by automated Edman degradation of overlapping peptides generated by a variety of selective cleavage procedures. The amino-terminus is blocked by a cyclized glutamine (pyroglutamic acid) residue, and the sequence of this region of the molecule was determined by mass spectrometry. Fibrolase is composed of 203 residues in a single polypeptide chain with a molecular weight of 22,891, as determined by the sequence. Its sequence is homologous to the sequence of the hemorrhagic toxin Ht-d of *Crotalus atrox* venom and with the sequences of two metalloproteinases from *Trimeresurus flavoviridis* venom. Microheterogeneity in the sequence was found at both the amino-terminus and at residues 189 and 192. All six cysteine residues in fibrolase are involved in disulfide bonds. A disulfide bond between cysteine-118 and cysteine-198 has been established and bonds between cysteines-158/165 and between cysteines-160/192 are inferred from the homology to Ht-d. Secondary structure prediction reveals a very low percentage of $\alpha$-helix (4%), but much greater $\beta$-structure (39.5%). Analysis of the sequence reveals the absence of asparagine-linked glycosylation sites defined by the consensus sequence: asparagine-X-serine/threonine.

**Keywords:** amino acid sequence; secondary structure; snake venom fibrolase

Fibrinolytic activity has been found in the venom of snakes from the families Crotalidae, Viperidae, and Elapidae, with venom from members of the Crotalidae having the highest levels of fibrinolytic activity (for a recent review, see Markland, 1988). Fibrolase is a direct-acting fibrinolytic enzyme from southern copperhead (*Agkistrodon contortrix contortrix*) venom that does not require blood-borne cofactors for activity (Guan et al., 1991). As previously described (Markland, 1983; Markland et al., 1988; Retzios & Markland, 1988), the enzyme is a nonhemorrhagic, zinc metalloproteinase which cleaves primarily the $\alpha$-chain of human fibrinogen and fibrin. Specific cleavage sites have been determined for several venom metalloproteinases using natural or synthetic substrates (Tu et al., 1981; Hagihara et al., 1985; Fox et al., 1986; Mori et al., 1987). For the venom fibrinolytic enzymes (Retzios & Markland, 1988; Willis & Tu, 1988), as well as the hemorrhagic metalloproteinases (Bjarnason & Fox, 1988–1989), cleavage appears to occur at the amino-terminal side of hydrophobic amino acids. In view of the clinical potential of venom fibrinolytic enzymes as thrombolytic agents (Markland et al., 1989; Willis et al., 1989), we undertook the sequence analysis of fibrolase to aid in both the cloning of the gene for this enzyme and the study of its function. The present communication describes the determination of the complete amino acid sequence of

Reprint requests to: Anne Randolph, Chiron Research Laboratories, Chiron Corporation, 4560 Horton Street, Emeryville, California 94608.

*Abbreviations:* CNBr, cyanogen bromide; TFA, trifluoroacetic acid; HPLC, high-performance liquid chromatography; PE-fibrolase, *S*-$\beta$-4-pyridylethyl-cysteine fibrolase; IBA, iodosobenzoic acid; SDS, sodium dodecyl sulfate; PAGE, polyacrylamide gel electrophoresis; DTNB, 5,5′-dithiobis(2-nitrobenzoic acid); Ht-d, *Crotalus atrox* hemorrhagic toxin d; TPCK, *N*-tosyl-L-phenylalanine chloromethyl ketone; FAB, fast atom bombardment.

fibrolase and represents the first report to date of the sequence of a nonhemorrhagic, fibrinolytic snake venom metalloproteinase.

## Results

The primary structure of fibrolase was determined by Edman degradation of overlapping fragments generated by a variety of selective cleavage procedures. Figure 1 shows the complete sequence of the 203 residues of fibrolase, with peptides important to the determination indicated below the sequence. The number of residues determined from a given peptide is shown by the length of the corresponding line under the sequence. The amino acid composition calculated for the complete sequence agrees with the composition determined by hydrolysis of the protein, as shown in Table 1. No data were obtained from Edman degradation of intact fibrolase, indicating that the amino-terminus is blocked either by a physiological mechanism or as an artifact of purification. Therefore, mass spectrometry of the amino-terminal CNBr fragment was employed to establish the structure of this region of the molecule.

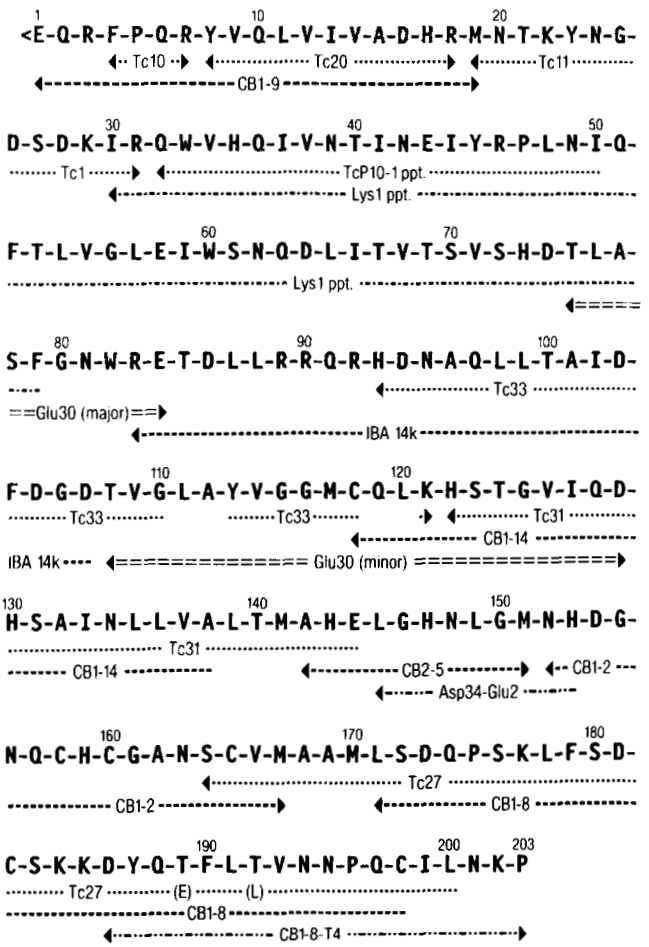


**Fig. 1.** The complete amino acid sequence of fibrolase. Peptides important in the sequence determination are indicated; the number of residues determined from a given peptide is shown by the length of the corresponding line under the sequence. There is heterogeneity in the sequence at positions 189 and 192; the residues found in CB1-8 are shown in the main sequence, with the residues found in Tc27 in parentheses below.

**Table 1.** *Amino acid composition of fibrolase*[a]

| Amino acid | Experimental value | Theoretical value |
|---|---|---|
| Asp + Asn | 28–31 | 31 |
| Glu + Gln | 19–20 | 20 |
| Ser | 11–12 | 12 |
| Gly | 12 | 12 |
| His | 8–9 | 10 |
| Arg | 9 | 9 |
| Thr | 12–13 | 13 |
| Ala | 11 | 11 |
| Pro | 6–7 | 5 |
| Tyr | 5–6 | 5 |
| Val | 12 | 14 |
| Met | 5 | 6 |
| Cys[b] | 7 | 6 |
| Ile | 9–10 | 12 |
| Leu | 20–21 | 21 |
| Phe | 6–7 | 6 |
| Lys | 7–8 | 7 |
| Trp[c] | ND | 3 |
| Total | | 203 |

[a] Results are expressed as residues per mole using the standard three letter code for the amino acids.
[b] Cysteine was determined as the *S*-β-4-pyridylethyl derivative.
[c] Tryptophan was not determined (ND) in the analysis.

### *CNBr peptides*

The peptides derived from CNBr cleavage of PE-fibrolase were fractionated by size on a Bio-Gel P-10 column in 50% acetic acid. The four pools obtained (data not shown) were dried in a Speed Vac and stored at −20 °C until use.

The first P-10 pool (CB1) was rechromatographed by reverse-phase HPLC using a C4 column; the chromatogram is shown in Figure 2. Automated Edman degradation of the pools yielded the complete sequence of CB1-2 and most of the sequence of CB1-8 and CB1-14. CB1-8 was subdigested with trypsin, and the peptides were separated by HPLC using a C18 column (data not shown). Sequence analysis of one of the resulting peaks (CB1-8-T4) resulted in the determination of the complete sequence of CB1-8.

CB1-9, which contains the blocked amino-terminal CB peptide, was sequenced by mass spectrometry, using FAB techniques. Figure 3 depicts the fragmentation pattern for CB1-9. These data indicate that the amino-terminus is blocked by a pyroglutamic acid residue (<Glu) and place the first 19 residues in the sequence. CB1-10, the HPLC peak immediately following CB1-9, contains the amino-

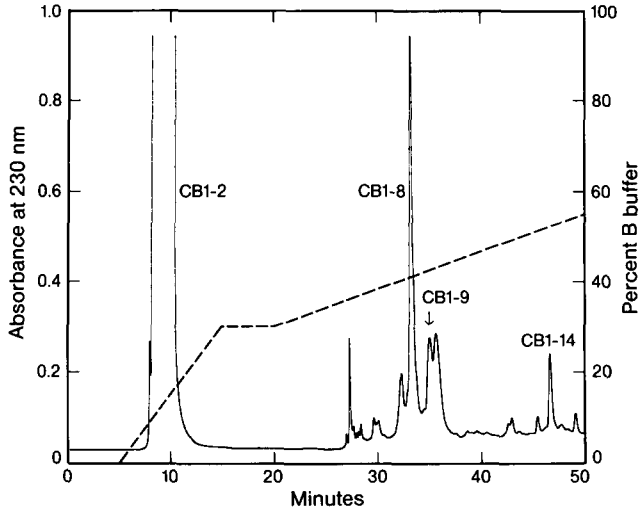


**Fig. 2.** HPLC separation of CB1. The sample (pool 1 from P-10 column of the CNBr fragments of PE-fibrolase) was loaded in 50% (v/v) acetic acid onto a C4 column (0.46 × 25 cm) equilibrated in 0.1% TFA. Peptides were eluted with an increasing gradient of 0.1% TFA in 60% *n*-propanol at a flow rate of 0.4 mL/min. The effluent was continuously monitored at 230 nm, and peaks were collected manually for composition and sequence analysis. The peptides present in peaks that yielded sequence information are indicated in the figure.

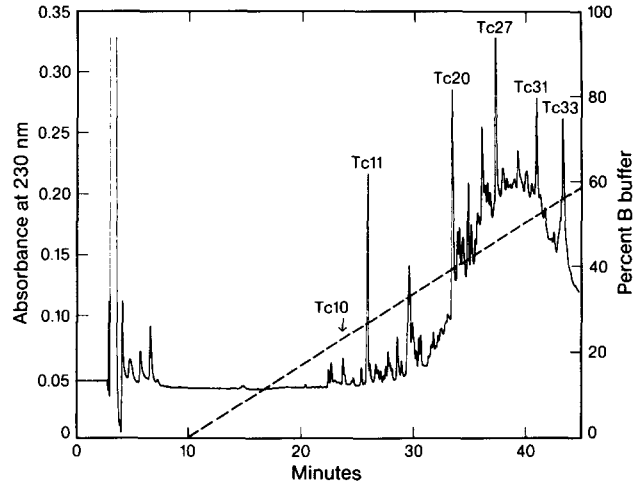


**Fig. 4.** HPLC separation of the tryptic digest of citraconylated PE-fibrolase. The digest was loaded onto a C18 column in 5 M urea. Peptides were eluted with an increasing gradient of 0.1% TFA in 80% acetonitrile. The flow rate was 0.8 mL/min. The effluent was continuously monitored at 230 nm, and peaks were collected manually for composition and sequence analysis. The peptides present in peaks that yielded sequence information are indicated in the figure.

terminus as determined by amino acid analysis. Although this fraction was not subjected to mass spectrometry, presumably it contains the amino-terminal CNBr peptide beginning at residue 2 (see *Lys-C peptides* below). Rechromatography of the second P-10 pool (CB2) using a C4 reverse-phase column resulted in one major and one minor peak (data not shown). The complete sequence of the major peak (CB2-5) was determined by Edman degradation.

### *Tryptic peptides*

Peptides resulting from the trypsin digestion of citraconylated PE-fibrolase were separated both by reverse-phase HPLC and gel filtration chromatography. The HPLC profile is shown in Figure 4, and several fractions were subjected to Edman degradation. Complete sequences were obtained for Tc10, Tc11, and Tc20; the larger peptides (Tc27, Tc31, and Tc33) were partially sequenced. Gel filtration of the digest resulted in four pools (data not shown). When the first peak (TcP10-1) was dissolved in 0.1% TFA, a pellet was obtained (TcP10-1 precipitate [ppt.]), which was partially sequenced by Edman degradation. We were able to obtain sequence information for TcP10-1 ppt., which begins with a glutamine residue, as a small fraction of the molecules were not blocked at the N-terminus. All but the first three residues of the amino-terminal region were confirmed by the sequence of the tryptic peptides Tc10, Tc11, and Tc20. The overlap for these peptides is provided by the mass spectrometry data.

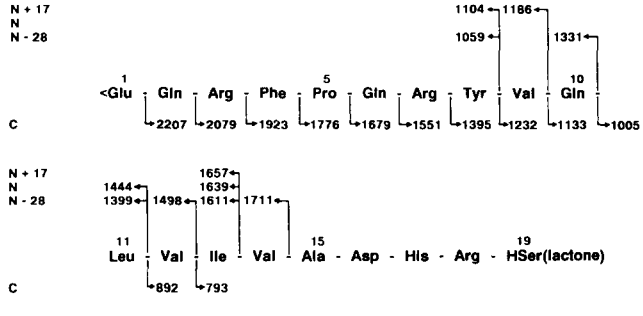


**Fig. 3.** Mass fragmentation pattern for peptide CB1-9. The amino- and carboxyl-terminal sequence ions observed in the spectrum are indicated above and below the sequence, respectively.

### *Lys-C peptides*

Gel filtration of the lys-C digest resulted in five pools (data not shown). The pellet obtained when the first pool was dissolved in 0.1% TFA (Lys1 ppt.) was not sequenced to its C-terminus. The second pool (Lys2) was rechromatographed on a C18 reverse-phase HPLC column (data not shown). The major peak (a doublet, Lys2-15) was analyzed by mass spectrometry and found to contain a mixture of two amino-terminal peptides differing in mass by one residue. The peak was asymmetric, and mass analysis showed two molecular ions: a strong signal at mass 2,580 and a weaker signal at mass 2,708. The molecular ion masses predict structures corresponding to the peptides

derived from (1) residues <Glu-2 through Lys-22 and (2) residues <Glu-1–Gln-2 through Lys-22, respectively (see Fig. 1).

### *Glu-C peptides*

Peptides resulting from cleavage of the citraconylated PE-fibrolase with endoproteinase glu-C were purified on a C18 reverse-phase HPLC column (Fig. 5). Only one of the many peaks was subjected to sequence analysis. Fraction Glu30 was found to contain two peptides by virtue of the presence of two residues at each of the first 10 cycles of degradation. The two sequences (labeled major and minor in Fig. 1) were easily distinguished, as they were present in very different molar amounts in the mixture.

### *Asp peptides*

The peptides of PE-fibrolase generated by acid cleavage were fractionated on a C18 HPLC column. Many peaks were obtained (data not shown), but only information derived from Asp34 was useful in the determination of the complete sequence of fibrolase. The amino acid composition of fraction Asp34 (5 Asp, 3 Glu, 3 Ser, 6 Gly, 4 His, 1 Arg, 2 Thr, 5 Ala, 1 Tyr, 3 Val, 1 Met, 1 Ile, 6 Leu, and 1 Lys) suggested that it contained a peptide spanning the region across the junction between CB2-5 and CB1-2. Although clearly not pure, this fraction was succinylated to block peptide amino-termini and digested with endoproteinase glu-C. The subdigest was sequenced directly to provide an important overlap between CB2-5 and CB1-2 (Fig. 1).

### *IBA peptides*

Cleavage after tryptophan residues in fibrolase is predicted to yield four peptides with molecular weights of 4,093, 3,224, 2,394, and 13,263 Da (Fig. 1). The fragments produced by reaction of PE-fibrolase with IBA were separated by SDS-PAGE; digestion was incomplete, and three bands at molecular weights of 23,000, 17,000, and 14,000 were clearly discernible after visualization with Coomassie blue. The large bands represent intact fibrolase and the carboxyl-terminal portion of the molecule, derived from incomplete cleavages after the second and third tryptophan residues, respectively. The 14,000-Da product was excised from a preparative gel, electroeluted, and purified by acetone precipitation. Sequence analysis provided the final overlap for the region of the molecule following Trp-82 (Fig. 1).

### *Th peptides*

In an attempt to locate the positions of the three disulfide bonds of fibrolase, the native molecule was cleaved with thermolysin. Fibrolase with intact disulfide bonds was very resistant to cleavage, and prolonged digestion in 8 M urea at elevated temperatures (60 °C) was necessary to cleave the protein. Peptides greater than seven residues (after dialysis) were separated on a reverse-phase C18 column. Peaks were collected (Fig. 6) and subjected to amino acid analysis. Fraction Th26 contained 2 Gly, 1 Met, 1 Cys, 2 Glu, 1 Val, 2 Asp, and 1 Pro (Cys was

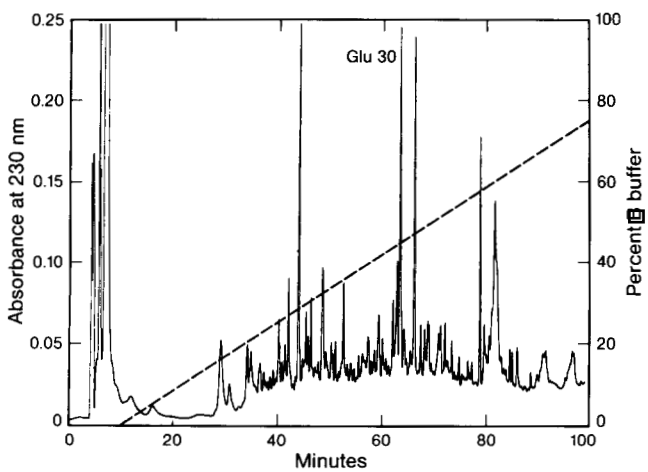


**Fig. 5.** HPLC separation of the endoproteinase glu-C digest of PE-fibrolase. The digest, after decitraconylation at pH 2, was loaded directly onto a C18 column equilibrated in 0.1% TFA. All conditions were as described in the legend to Figure 4. The peptide present in the peak that yielded sequence information is indicated in the figure.

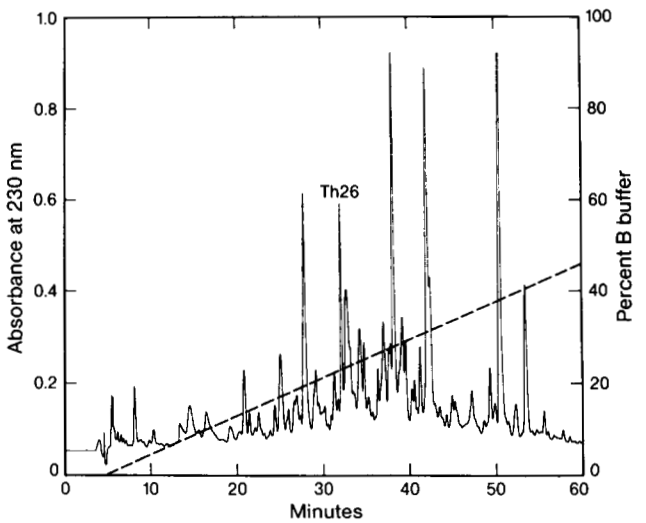


**Fig. 6.** HPLC separation of the thermolysin digest of native fibrolase. The digest was dialyzed to remove small peptides (<7 residues) and loaded onto a C18 column equilibrated in 0.1% TFA. All conditions were as described in the legend to Figure 4. Fraction Th26 was found to contain peptides involving one of the three disulfide bonds in the molecule.

neither derivatized nor oxidized, and recovery was not quantitative), corresponding to a mixture of two peptides: Gly-115–Gly-Met-Cys-Gln-119 and Val-193–Asn-Asn-Pro-Gln-Cys-198. The results indicate that cysteine residue 118 is connected to cysteine residue 198 by a disulfide bond in native fibrolase (Fig. 1).

### *Heterogeneity of sequence*

One tryptic peptide (Tc27) was the result of an unexpected cleavage between Asn-163 and Ser-164. Sequence analysis of this peptide clearly revealed a glutamic acid residue at position 26 and a leucine residue at position 29 (shown in parentheses under the sequence in Fig. 1 at positions 189 and 192, respectively). Sequencing of the corresponding CNBr peptide (CB1-8), however, clearly revealed threonine residues at both of these positions, with no glutamic acid or leucine at the appropriate cycle. Also, no indication of threonine residues at these positions was found upon careful analysis of the data for Tc27. Therefore, it appears that there is microheterogeneity at these positions in the fibrolase preparation.

## Discussion

### *General features of the sequence*

The complete primary structure of fibrolase comprises 203 amino acids, with microheterogeneity at the N-terminus and at two internal sites. The amino acid composition calculated for the complete sequence agrees with the composition determined by hydrolysis of the protein; the molecular weight derived from the sequence is 22,891 (average mass calculation), with a theoretical pI of 6.5. Inspection of the sequence reveals the absence of consensus sites for *N*-linked glycosylation (Asn-X-Ser/Thr). Separate studies have revealed that the pure protein, when analyzed using sensitive carbohydrate stains (Dubray & Bezard, 1982) after native PAGE, contains no glycosylation of any type.

The results from direct Edman degradation of fibrolase and CNBr, trypsin, and lys-C cleavages indicate the amino-terminus of the enzyme is blocked by cyclization of a glutamine residue. In addition, there is heterogeneity at the amino-terminus, with some molecules beginning with <Glu-Gln-Arg-Phe... and some beginning with <Glu-Arg-Phe.... The use of mass spectrometry made the evaluation of the blockage simple and direct, without the need to subject the protein or peptides to enzymatic hydrolysis with pyroglutamylaminopeptidase.

Sequence analysis has revealed the presence of microheterogeneity in the fibrolase preparation at internal sites as well. This heterogeneity is probably responsible for the appearance of the two bands observed with immobilized pH gradient isoelectric focusing (Guan & Markland, 1988). Because the difference involves the substitution of threonine for glutamic acid at position 189, it should explain the isoelectric point heterogeneity, as there is an absence of glycosylation. The other substitution, threonine for leucine at position 192, would produce no change in the isoelectric point of the protein. Because the protein was purified from a venom pool of several individual snakes, the heterogeneity may be due to differences in fibrolase among individual snakes. Alternatively, there could be more than one fibrolase isozyme present in a given snake. To address this issue, a study involving the characterization of the protein from the venom of individual snakes is in progress (F.S. Markland & N. Egen, unpubl.).

Titration of free thiol groups in native fibrolase with DTNB revealed the absence of free sulfhydryl groups in the molecule, indicating that the six cysteine residues are involved in three disulfide bonds. Thermolysin digestion of intact fibrolase yielded only one product that could be ascribed to a cystine peptide, resulting from a disulfide bond between Cys-118 and Cys-198. We were unable to locate fragments that would provide information to describe the positions of the two other disulfide bonds in the molecule. It is possible that those regions are resistant to proteolysis even under the stringent conditions used in this study.

### *Secondary structure prediction*

The method of Chou and Fasman (1978) was used for the prediction of the secondary structure of fibrolase. The predicted structure is shown in Figure 7, and its examination reveals a very low percentage of alpha-helical regions (only 4% of the sequence). However, there is a strikingly higher amount of beta-structure in the enzyme (39.5% of the sequence). The same low percentage of al-

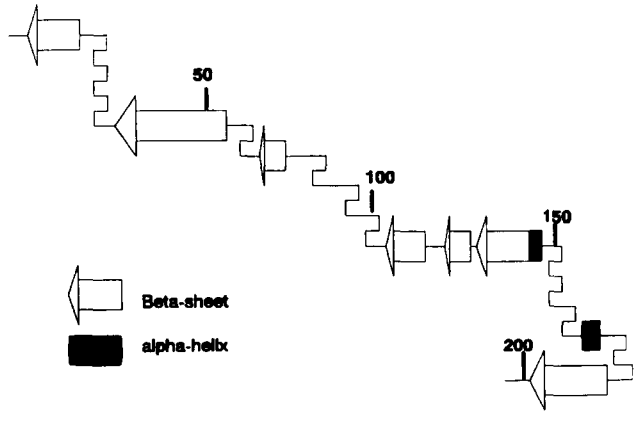


**Fig. 7.** Prediction of secondary structure of fibrolase using the method of Chou and Fasman (1978). As shown in the figure, the amino-terminus of the polypeptide is at the upper left and extends down to the carboxyl-terminus at the lower right. This method predicts only 4% $\alpha$-helix and 39.5% $\beta$-structure for fibrolase.

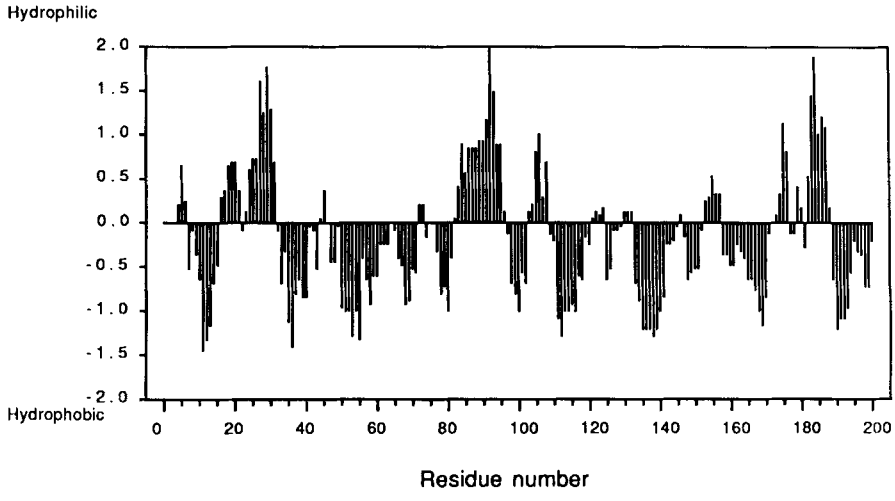


**Fig. 8.** Hydrophobicity plot for fibrolase using the method of Hopp and Woods (1981). Three hydrophilic domains are predicted for the N-terminus, middle, and C-terminus of the molecule.

pha-helical structure is also predicted from the sequences of HR2a and $H_2$-proteinase. The hydrophobicity plot derived from the scale of Hopp and Woods (1981) is depicted in Figure 8; the structure is characterized by a striking preponderance of hydrophobic regions, separated into three distinct areas by hydrophilic stretches at the N-terminus, middle, and C-terminus. X-ray structure determination will not only serve to clarify the interesting arrangement of these regions, which has not been previously reported, but also to define the zinc-binding site with its coordination residues.

### *Homology considerations*

Unexpectedly, the amino acid sequence of fibrolase is homologous to the hemorrhagic metalloproteinase Ht-d from *Crotalus atrox* (Shannon et al., 1989). Both molecules have 203 residues, and the alignment of the two sequences is shown in Figure 9. Comparison of the sequences reveals 120 identical residues out of 203 (59% identity). Notably, Ht-d, like fibrolase, was found to be heterogeneous at the amino-terminus with some of the molecules beginning at residue 1 and some at residue 2 (Shannon et al., 1989). It is interesting to note that, despite the significant structural homology between fibrolase and the hemorrhagic metalloproteinase Ht-d, the fibrinolytic enzyme almost completely lacks hemorrhagic activity (A.L. Guan, A.D. Retzios, & F.S. Markland, unpubl.). This observation raises an interesting question concerning the structural parameters that determine the divergent biological activities of the two proteins. Resolution of the question concerning structural features that distinguish the fibrinolytic enzymes from hemorrhagic metalloproteinases is not readily apparent from a consideration of their primary structures, however, and must await X-ray structure determinations.

The four cysteine residues in Ht-d are conserved in fibrolase; fibrolase has two additional Cys residues that are not present in Ht-d (Fig. 9). Ht-d contains two disulfide bonds, one between Cys-118 and Cys-198 (the two residues found to be linked in fibrolase) and one between

```
                                10                                  20
F       <E-Q-R-F-P-Q-R-Y-V-Q-L-V-I-V-A-D-H-R-M-N-T-K-Y-N-G-
H            N L         I E     V           V F M       S

                30                                  40                  50
F       D-S-D-K-I-R-Q-W-V-H-Q-I-V-N-T-I-N-E-I-Y-R-P-L-N-I-Q-
H         L N T     T R     E       F     G F     S       H

                                60                                  70
F       F-T-L-V-G-L-E-I-W-S-N-Q-D-L-I-T-V-T-S-V-S-H-D-T-L-A-
H       V S   T D                   Q   N I Q   A   S       N

            80                              90                          100
F       S-F-G-N-W-R-E-T-D-L-L-R-R-Q-R-H-D-N-A-Q-L-L-T-A-I-D-
H       A   A E               N   K S                           E

                    110                             120
F       F-D-G-D-T-V-G-L-A-Y-V-G-G-M-C-Q-L-K-H-S-T-G-V-I-Q-D-
H       L   E E   L       P L   T     D P   L   I   I V

            130                         140     *       *       150
F       H-S-A-I-N-L-L-V-A-L-T-M-A-H-E-L-G-H-N-L-G-M-N-H-D-G-
H           P         M G V                                 E

                160                             170                     180
F       N-Q-C-H-C-G-A-N-S-C-V-M-A-A-M-L-S-D-Q-P-S-K-L-F-S-D-
H       K D   L R     S L   I   R P G   T K G R   Y E

                                190                             200
F       C-S-K-K-D-Y-Q-T-F-L-T-V-N-N-P-Q-C-I-L-N-K-P
H       D   M H Y   E R     K Q Y K
```

**Fig. 9.** Comparison of the amino acid sequences of fibrolase and the hemorrhagic toxin Ht-d from *Crotalus atrox*. The complete sequence of fibrolase (labeled F) is shown with the sequence of Ht-d (labeled H) given below only at positions where the two sequences differ. The six cysteine residues in fibrolase are highlighted, and the putative zinc-binding ligands are indicated by asterisks (*).

Cys-158 and Cys-165. Presumably, in fibrolase, Cys-158 is linked to Cys-165 and the remaining two Cys residues, Cys-160 and Cys-192, are linked together.

It is of interest to note that fibrolase is also homologous to two recently sequenced metalloproteinases from *Trimeresurus flavoviridis* venom, HR2a (Miyata et al., 1989) and $H_2$-proteinase (Takeya et al., 1989), of 202 and 201 residues, respectively. Fibrolase displays 59–61% sequence homology with the two proteins; the Pearson and Lipman (1988) algorithm for homology assessment awarded scores of 670–683 out of a maximum of 1,200 for complete identity. The positions of the disulfide bonds in these two *T. flavoviridis* metalloproteinases support the observation of the disulfide bond between Cys-118 and Cys-198 in fibrolase, but the other two disulfide bonds in HR2a are located between Cys-161/Cys-164 and Cys-159/Cys-181 (the number system is slightly different due to the shorter length of HR2a). Thus, there appears to be a difference in the definition of the structures determined in the two different laboratories.

Fibrolase (Markland et al., 1988; Retzios & Markland, 1988), Ht-d (Shannon et al., 1989), and the two enzymes from *T. flavoviridis* venom (Miyata et al., 1989; Takeya et al., 1989) are zinc metalloproteinases that show complete identity in the region surrounding the putative zinc-binding site, which is inferred by homology to thermolysin (Matthews et al., 1972). Figure 10 presents a comparison of the zinc-binding region in fibrolase, Ht-d, HR2a, $H_2$-proteinase, thermolysin, human fibroblast collagenase (Goldberg et al., 1986), *Serratia* protease (Nakahama et al., 1986), and *Bacillus subtilis* neutral protease (Vasantha et al., 1984). There is a high degree of homology at this site among all of the metalloproteinases in this example. The two putative zinc ligands (His-143 and His-147 in fibrolase) are completely conserved as well as a Glu residue following the first His and a hydrophobic residue (Leu or Val) two residues after the second His. In addition, there is homology among the majority of the eight proteins at five other positions in this region. Regions of this type have been proposed as a consensus region for recognizing members of the metalloproteinase family (McKerrow, 1987; Jongeneel et al., 1989). However, fibrolase and other venom-derived metalloproteinases do not appear to share significant additional sequence homology with other, well-characterized neutral metalloendopeptidases.

A search for homologous proteins in the NBRF protein database using the Pearson and Lipman (1988) algorithm (FASTA) revealed that the highest alignment score (96 out of a possible 1,200) was obtained with the hypothetical BRRF1 protein from the Epstein-Barr virus, or with a portion of the env protein of the human immunodeficiency virus (score = 75). Homology scores with human collagenase (54) and with the metalloproteinase of *Serratia* (60) were considerably lower. Interestingly, optimized alignment of fibrolase and BRRF1 protein (Baer et al., 1984) shows that the viral protein retains the essential components of the zinc-binding site, including the consensus signature. The significance, if any, of this identity remains to be established, although it raises some intriguing questions concerning symbiotic relationships between viruses and snakes.

## Materials and methods

### *Materials*

Fibrolase was purified from southern copperhead (*A. c. contortrix*) venom (Biotoxins, Inc., St. Cloud, Florida) using the four-step procedure described previously (Markland et al., 1988). Other enzymes and chemicals used in this work were obtained as follows: CNBr, constant boiling HCl, ninhydrin, and IBA were products of Pierce Chemical Co., Rockford, Illinois; all sequencer reagents were obtained from Applied Biosystems, Foster City, California; 4-vinyl pyridine, *N*-ethyl morpholine, citraconic anhydride, pyridine, and *p*-cresol were produced by Aldrich Chemical Co., Milwaukee, Wisconsin; trypsin-TPCK was purchased from Sigma Chemical Co., St. Louis, Missouri; formic acid (95–97%) was purchased from Kodak Chemical Co., Rochester, New York; endoproteinase lys-C, endoproteinase glu-C, and thermolysin

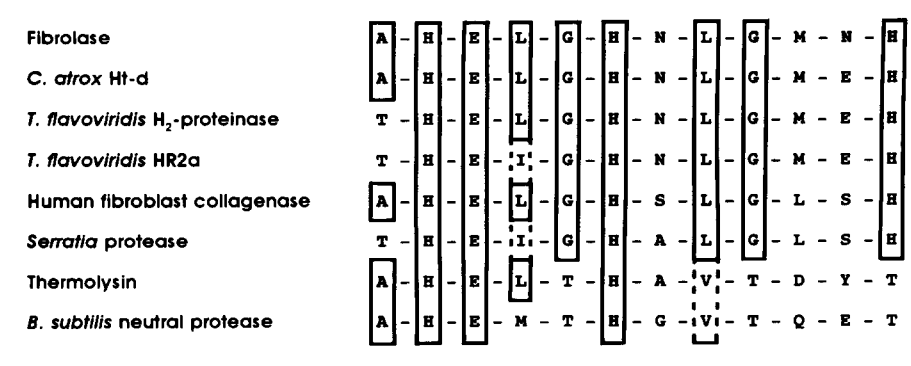


**Fig. 10.** Comparison of zinc-binding regions. The zinc-binding regions of fibrolase, *C. atrox* Ht-d, two *T. flavoviridis* metalloproteinases, thermolysin, human fibroblast collagenase, *Serratia* protease, and *B. subtilis* neutral protease are shown. The zinc-binding ligands are indicated by asterisks (*). Residues are boxed where homology is seen among more than two of the molecules. Identical residues are indicated by solid boxes and dashed boxes indicate conservative amino acid substitutions.

were obtained from Boehringer Mannheim Biochemicals, Indianapolis, Indiana. Bio-Gel media were obtained from Bio-Rad Laboratories, Hercules, California; fluorescamine was a product of Roche, Nutley, New Jersey. Electrophoresis purity reagent SDS purchased from Bio-Rad was recrystallized from ethanol–water as described by Hunkapiller et al. (1983). *N*-ethylmorpholine and pyridine were distilled over ninhydrin (1 g/L). All other reagents were of the highest purity commercially available. Reverse-phase HPLC columns (C4 and C18) were purchased from Vydac, Hesperia, California.

TFA solutions were prepared on a w/v basis, and all acetic acid solutions were prepared on a v/v basis.

### *Amino acid analysis*

Amino acid compositions were obtained as described by Bidlingmeyer et al. (1984) with methanol substituted for ethanol in the derivatization mixture (Cohen et al., 1986). Analyses were performed on protein samples hydrolyzed in the vapors of constant-boiling HCl containing 1% (v/v) phenol for 22–24 h, in vacuo, at 110 °C. Phenylthiocarbamyl-amino acids were separated by reverse-phase HPLC using a Waters 840 chromatography system (Pico-Tag method; Waters Associates, Milford, Massachusetts).

### *Sequence analysis*

Automated Edman degradation was performed on an Applied Biosystems Model 470A gas-phase protein sequencer equipped with a 120A on-line phenylthiohydantoin-amino acid analyzer (Applied Biosystems, Foster City, California). All sequencer runs were made using 03RPTH cycles, supplied by the manufacturer.

### *Mass spectrometry*

Peptide samples were sent to M-Scan Ltd. (Berkshire, England) for analysis by mass spectrometry. The samples were dissolved in 5 μL of 5% aqueous acetic acid and 2-μL aliquots were loaded onto a target previously smeared with 2–4 μL of glycerol. FAB analysis was carried out on a VG Analytical ZAB high field mass spectrometer operating at an accelerating voltage of 8 kV for a 300 mass range at full sensitivity. A CsI calibration spectrum was run as well. An M-Scan FAB gun was used to generate spectra, using xenon gas, operating at 10 kV at a beam strength of 15 μA.

### *Reaction of fibrolase with DTNB*

The amount of free sulfhydryl groups in native fibrolase was quantitated with DTNB according to the procedure of Ellman (1959).

### *Reduction and alkylation*

Fibrolase (0.5 mg in 100 μL of 0.2 M *N*-ethylmorpholine acetate [pH 8.7] containing 6 M guanidine-HCl and 3 mM EDTA [reduction buffer]) was reduced with 1 μL of 2-mercaptoethanol at 50 °C under argon for 1 h. Following reduction, 10 μL of 4-vinyl pyridine (Friedman et al., 1970) were added under argon and allowed to react for 2 h at room temperature. PE-fibrolase was dialyzed against 5% acetic acid and stored at −20 °C until use.

### *CNBr cleavage*

PE-fibrolase was dissolved in 70% formic acid (v/v) (4.5 mg protein/mL), and crystalline CNBr was added (1:1 ratio to protein by weight). The reaction was allowed to proceed under argon in the dark at room temperature. An additional aliquot of CNBr was added to the reaction mixture after 24 h. After 36 h, the reaction mixture was dried in a Speed Vac (Savant Instruments Inc., Farmingdale, New York). The resulting peptides were dissolved in 280 μL of 50% acetic acid and loaded on a Bio-Gel P-10 (200–400 mesh) column as described below (see *Peptide separations*). One CNBr peptide (CB1-8) was subdigested with trypsin. The peptide was dissolved (1.5 μg/100 μL final concentration) in 0.1 M ammonium bicarbonate containing 2% (w/w) trypsin-TPCK, and digestion was allowed to proceed at room temperature overnight.

### *Trypsin digestion*

Fibrolase was citraconylated prior to digestion with trypsin in order to restrict cleavages to arginine residues. PE-fibrolase was dissolved in reduction buffer (~5 mg/mL), and citraconic anhydride was added (5 × 5-μL portions per 1 mg of fibrolase) while the pH of the solution was maintained between 8 and 9 with 1 N NaOH. After the pH was stabilized at 8.5, the mixture was dialyzed against 50 mM *N*-ethylmorpholine acetate (pH 8.5) and dried in a Speed Vac. The citraconylated protein was dissolved in 0.5 mL of 0.2 M Tris-HCl (pH 8.4) containing 2 M urea and 0.1 mM $CaCl_2$ and 2% (w/w) trypsin-TPCK was added. The reaction was allowed to proceed overnight at room temperature. The digest was titrated to pH 2 with TFA and left at room temperature overnight to decitraconylate. The resulting peptide mixture was cloudy and was clarified by increasing the urea concentration to 5 M prior to separating the peptides using HPLC or dried in a Speed Vac and dissolved in a minimum volume of 50% acetic acid for loading on a Bio-Gel P-10 column.

### *Endoproteinase lys-C digestion*

PE-fibrolase was dissolved in 0.3 M Tris-HCl (pH 8.6) containing 0.1 mM $CaCl_2$ and 5.0 M urea (2.3 mg fibro-

lase/3 mL) and 4% (w/w) endoproteinase lys-C was added. Following digestion at room temperature for 24 h, the sample was dialyzed against 50% acetic acid in Spectra/Por 3 (Spectrum Medical Industries, Inc., Los Angeles, California) dialysis tubing that had been boiled for 5 min in 0.1 M EDTA (pH 8.0) to reduce the pore size so that peptides of seven residues or more are retained (Fowler, 1978). The digest was dried in a Speed Vac and redissolved in a small volume of 50% acetic acid prior to the initial peptide separation using Bio-Gel P-30.

*Endoproteinase glu-C digestion*

PE-fibrolase was citraconylated as described above under *Trypsin digestion* to increase its solubility, dialyzed into 50 mM *N*-ethylmorpholine acetate (pH 8.0), and partially dried to a concentration of 8 mg fibrolase/mL. Endoproteinase glu-C was added (2% w/w) and the mixture was incubated at 37 °C. After 24 h, a second aliquot (2%) of enzyme was added, and after 40 h, the reaction was stopped and the protein decitraconylated by titration to pH 2 with TFA and incubation at room temperature for at least 3 h.

*Cleavage at tryptophan residues with IBA*

Fibrolase was cleaved at tryptophan residues as described by Fontana et al. (1983). PE-fibrolase was dissolved in 80% acetic acid, 4 M guanidine-HCl (0.6 mg/120 μL), and IBA (1.2 mg) and 1 μL *p*-cresol were added. After reaction for 24 h in the dark, the sample was diluted to 0.02 M guanidine-HCl and precipitated by the addition of 1/9 volume of 100% trichloroacetic acid containing 4 mg/mL sodium deoxycholate. The resulting pellet was washed with cold acetone to remove the sodium deoxycholate, trichloroacetic acid, and any remaining IBA.

*Acid cleavage in TFA*

PE-fibrolase was dissolved in 0.1% TFA (0.35 mg fibrolase/mL) and heated at 100 °C for 1 h. One peptide resulting from this cleavage (Asp34) was subdigested with endoproteinase glu-C. First, the peptide was dissolved in reduction buffer that had been titrated to pH 7.0, and succinic anhydride (>100-fold molar excess over lysine residues) was added while the pH of the solution was maintained between 7 and 8 with 1 N NaOH. After the pH was stabilized at 7, the mixture was dialyzed against 0.1 M ammonium bicarbonate and dried in a Speed Vac. The succinylated peptide was dissolved in 0.1 M ammonium acetate (pH 4.0), and 4% (w/w) endoproteinase glu-C was added. Following reaction for 4 h at 37 °C, the digest was subjected to Edman degradation.

*Cleavage of native fibrolase with thermolysin*

Native fibrolase (2.5 mg with disulfide bonds intact) was dissolved in 250 μL of 0.1 M pyridine acetate (pH 6.5) containing 8 M urea and 2 mM $CaCl_2$ and digested with 10% (w/w) thermolysin at 60 °C for 48 h. The digest was then dialyzed against water in tubing that retains peptides of seven or more residues (see *Endoproteinase lys-C digestion*).

*Peptide nomenclature*

Peptides resulting from cleavages of fibrolase with CNBr (CB), trypsin following citraconylation (Tc), endoproteinase lys-C (Lys), endoproteinase glu-C (Glu), TFA (Asp), and thermolysin (Th) are designated by the respective letters in parentheses followed by arabic numeral(s) indicating the peak elution order in the respective chromatograms. When a peptide pool was rechromatographed, a number indicating peak location in the second chromatogram is added to the designation after a hyphen (-). Because the total tryptic digest of citraconylated fibrolase was chromatographed on both HPLC and Bio-Gel P-10, the Bio-Gel pools have a P-10 added to their designation before the numeral to distinguish them from HPLC-derived pools. When a precipitate was formed following dissolution of a peptide pool, "ppt." is added following the designation of the insoluble fraction. Peptides resulting from the subdigestion of a fragment with trypsin (T) or endoproteinase glu-C (Glu) have the corresponding letter(s), followed by their order in the parent fragment, added to the peptide designation. Peptides resulting from cleavage of fibrolase with IBA are designated IBA, followed by the apparent molecular weight of the fragment on SDS-PAGE.

*Peptide separations*

The CNBr digest of PE-fibrolase (4.5 mg in 280 μL 50% acetic acid) was initially separated by gel filtration chromatography. The digest was loaded on a 0.7 × 46.5-cm column of Bio-Gel P-10 (200–400 mesh) in 50% acetic acid and eluted at a flow rate of 2.6 mL/h. Effluent fractions were monitored by reaction with fluorescamine (Bohlen et al., 1975), and the four pools obtained were dried in a Speed Vac. The first P-10 pool, CB1, was redissolved in 50% acetic acid and rechromatographed by reverse-phase HPLC using a C4 column (250 × 4.6 mm). The peptides were eluted with an increasing gradient of 60% *n*-propanol containing 0.1% TFA at a flow rate of 0.4 mL/min. The effluent was monitored at 230 nm, and the resulting peaks were collected for composition and sequence analysis. Similarly, CB2 was dissolved in 0.1% TFA and rechromatographed on a C4 column. The tryptic subdigest of CB1-8 was fractionated using a C18

column (250 × 4.6 mm) eluted with an increasing gradient of 80% acetonitrile containing 0.1% TFA at a flow rate of 0.8 mL/min, and the resulting peaks were collected for analysis.

Peptides resulting from the tryptic digest of citraconylated PE-fibrolase were fractionated both by reverse-phase HPLC using a C18 column as described above and by gel filtration chromatography on Bio-Gel P-10 as follows. The digest (2.4 mg in 400 μL of 50% acetic acid) was loaded on a 0.7 × 46-cm P-10 (200–400 mesh) column and eluted with 50% acetic acid at a flow rate of 2 mL/h. Effluent fractions were monitored by reaction with fluorescamine, and the pools were dried in a Speed Vac. The first pool (TcP10-1) was dissolved in 1.0 ml of 0.1% TFA and centrifuged in a microfuge for 15 min. The resulting pellet (TcP10-1 ppt.) was dissolved in 50% acetic acid and subjected to Edman degradation.

The endoproteinase lys-C digest of PE-fibrolase (2.3 mg in 300 μL of 50% acetic acid) was loaded on a 0.7 × 43-cm column of Bio-Gel P-30 (200–400 mesh). The column was eluted with 50% acetic acid at a flow rate of 2 mL/h, and fractions were monitored by reaction with fluorescamine. The five pools were dried in a Speed Vac and redissolved in 0.1% TFA. Pool 1 yielded a pellet following centrifugation (Lys1 ppt.) that was dissolved in 50% acetic acid and subjected to Edman degradation. Pool 2 was rechromatographed using a C18 column as described above.

Peptides resulting from the digestion of PE-fibrolase with endoproteinase glu-C, from acid cleavage of PE-fibrolase, and from digestion of native fibrolase with thermolysin were separated by reverse-phase HPLC using a C18 column as described above. The peaks obtained were collected for composition and sequence analysis.

Peptides resulting from the digestion of PE-fibrolase with IBA were separated using preparative SDS-PAGE as described by Hunkapiller et al. (1983). The fragments were resolved on a 15% polyacrylamide gel, and the bands were stained, excised, and electro-eluted as described (Hunkapiller et al., 1983). The electro-eluted samples were dried in a Speed Vac; SDS and most of the Coomassie blue were removed by ion pair extraction using solvent A as described by Konigsberg and Henderson (1983). The pellets were dissolved in a small volume of 0.1% SDS (less than 100 μg SDS) and subjected to Edman degradation.

### *Secondary structure prediction and homology with known sequences*

The method of Chou and Fasman (1978) was used for the prediction of the secondary structure of fibrolase. The Pearson and Lipman (1980) algorithm for homology assessment was used to compare sequences of fibrolase with several other venom metalloproteinases and with proteins in the NBRF protein database. A score of 1,200 is awarded by the method for full identity. The hydrophilicity scale of Hopp and Woods (1981) was used to develop a hydrophobicity plot averaged over 6 residues. The software package known as MacProMass (Lee & Vemuri, 1989) was used to generate the figure with a Macintosh computer and printer.

## Acknowledgments

The authors thank Drs. Pablo Valenzuela and Frank Marcus of Chiron Corporation for helpful discussions during the course of this work and Marion Merrell Dow, Incorporated, for its support and encouragement in the pursuit of clinical applications of the protein. Partial support was provided to F.S.M. by the National Heart, Lung and Blood Institute, grant HL 31389, and by Cortech, Incorporated, Denver, Colorado.

## References

Baer, R., Bankier, A.T., Biggin, M.D., Deininger, P.L., Farrell, P.J., Gibson, T.J., Hartfull, G., Hudson, G.S., Satchwell, S.C., Seguin, C., Tiffnell, P.S., & Barrell, B.G. (1984). DNA sequence and expression of B95-8 Epstein-Barr virus genome. *Nature 310*, 207–211.

Bidlingmeyer, B.A., Cohen, S.A., & Tarvin, T.L. (1984). Rapid analysis of amino acids using pre-column derivatization. *J. Chromatogr. 336*, 93–104.

Bjarnason, J.B. & Fox, J.W. (1988–1989). Hemorrhagic toxins from snake venoms. *Toxicol. Toxin Rev. 7*, 121–209.

Bohlen, P., Stein, S., Stove, J., & Udenfriend, S. (1975). Automatic monitoring of primary amines in preparative column effluents with fluorescamine. *Anal. Biochem. 67*, 438–444.

Chou, P.Y. & Fasman, G.S. (1978). Prediction of the secondary structure of proteins from their amino acid sequence. *Adv. Enzymol. 47*, 45–148.

Cohen, S.A., Bidlingmeyer, S.A., & Tarvin, T.L. (1986). PITC derivatives in amino acid analysis. *Nature 320*, 769–770.

Dubray, G. & Bezard, G. (1982). A highly sensitive periodic acid-silver stain for 1,2-diol groups of glycoproteins and polysaccharides in polyacrylamide gels. *Anal. Biochem. 119*, 325–329.

Ellman, G.L. (1959). Tissue sulfhydryl groups. *Arch. Biochem. Biophys. 82*, 70–77.

Fontana, A., Dalzoppo, D., Grandi, C., & Zambonin, M. (1983). Cleavage at tryptophan with o-iodosobenzoic acid. *Methods Enzymol. 91*, 311–317.

Fowler, A.V. (1978). Amino acid sequence of beta-galactosidase. *J. Biol. Chem. 253*, 5499–5504.

Fox, J.W., Campbell, R., Beggerly, L., & Bjarnason, J.B. (1986). Substrate specificities and inhibition of two hemorrhagic zinc proteases Ht-c and Ht-d from *Crotalus atrox* venom. *Eur. J. Biochem. 156*, 65–72.

Friedman, M., Krull, L.H., & Cavins, J.F. (1970). The chromatographic determination of cystine and cysteine residues in proteins as *S*-β-(4-pyridylethyl)cysteine. *J. Biol. Chem. 245*, 3868–3871.

Goldberg, G.I., Wikelm, S.M., Kronberger, A., Bquer, F.A., Grant, G.A., & Eisen, A.Z. (1986). Human fibroblast collagenase. *J. Biol. Chem. 261*, 6600–6605.

Guan, A.L. & Markland, F.S. (1988). Isoelectric focusing in immobilized pH gradients of a snake venom fibrinolytic enzyme. *J. Biochem. Biophys. Methods 16*, 215–226.

Guan, A.L., Retzios, A.D., Henderson, G.N., & Markland, F.S. (1991). Purification and characterization of a fibrinolytic enzyme from the venom of the southern copperhead snake (*Agkistrodon contortrix contortrix*). *Arch. Biochem. Biophys. 289*, 197–207.

Hagihara, S., Komori, Y., & Tu, A.T. (1985). Proteolytic specificity of hemorrhagic toxin b from *Crotalus atrox* (western diamondback rattlesnake) venom. *Comp. Biochem. Physiol. 82*, 21–27.

Hopp, T.P. & Woods, K.R. (1981). Prediction of protein antigenic determinants from amino acid sequences. *Proc. Natl. Acad. Sci. USA 78*, 3824–3828.

Hunkapiller, M.W., Lujan, E., Ostrander, F., & Hood, L.E. (1983). Isolation of microgram quantities of protein from polyacrylamide gels for amino acid sequence analysis. *Methods Enzymol. 91*, 227-236.

Jongeneel, C.V., Bouvier, J., & Bairoch, A. (1989). A unique signature identifies a family of zinc-dependent metalloproteinases. *FEBS Lett. 242*, 211-214.

Konigsberg, W.H. & Henderson, L. (1983). Removal of sodium dodecylsulfate from proteins by ion-pair extraction. *Methods Enzymol. 91*, 254-259.

Lee, T.D & Vemuri, S. (1989) MacProMass software. Beckman Research Institute of the City of Hope, Department of Immunology, Duarte, California.

Markland, F.S. (1983). Rattlesnake venom enzymes that interact with components of the hemostatic system. *J. Toxicol. Toxin Rev. 2*, 119-160.

Markland, F.S. (1988). Fibrin(ogen)olytic enzymes from snake venoms. In *Hemostasis and Animal Venoms* (Pirkle, H. & Markland, F.S., Eds.), pp. 149-172. Marcel Dekker, New York.

Markland, F.S., Bookstein, J.J., & Machado, T. (1989). *In vivo* fibrinolytic activity of fibrolase, a metalloproteinase from the venom of *Agkistrodon contortrix contortrix. Thromb. Haemostasis 62*, 121.

Markland, F.S., Reddy, K.N.N., & Guan, L. (1988). Purification and characterization of a direct-acting fibrinolytic enzyme from southern copperhead venom. In *Hemostasis and Animal Venoms* (Pirkle, H. & Markland, F.S., Eds.), pp. 173-189. Marcel Dekker, New York.

Matthews, B.W., Colman, P.M., Janonius, J.N., Titani, K., Walsh, K.A., & Neurath, H. (1972). Structure of thermolysin. *Nature New Biol. 238*, 41-43.

McKerrow, J.H. (1987). Human fibroblast collagenase contains an amino acid sequence homologous to zinc binding site of *Serratia* protease. *J. Biol. Chem. 262*, 5943.

Miyata, T., Takeya, H., Ozeki, Y., Arakawa, M., Tokunaga, F., Iwanaga, S., & Omori-Satoh, T. (1989). The primary structure of hemorrhagic protein, HR2a, isolated from the venom of *Trimeresurus flavoviridis. J. Biochem. 105*, 847-853.

Mori, N., Tashaiki, N., Sugihara, H., & Tu, A.T. (1987). Biochemical characterization of hemorrhagic toxins with fibrinogenase activity isolated from *Crotalus ruber ruber* venom. *Arch. Biochem. Biophys. 253*, 108-121.

Nakahama, K., Yoshimure, K., Marumoto, R., Kikuchi, I.S.L., Toshiharu, H., & Matsubara, H. (1986). Cloning and sequencing of *Serratia* protease gene. *Nucleic Acids Res. 14*, 5843-5855.

Pearson, W.R. & Lipman, D.J. (1988). Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. USA 85*, 2444-2448.

Retzios, A.D. & Markland, F.S. (1988). A direct-acting fibrinolytic enzyme from the venom of *Agkistrodon contortrix contortrix*: Effects of various components of the human blood coagulation and fibrinolytic systems. *Thromb. Res. 52*, 541-552.

Shannon, J.D., Baramova, E.N., Bjarnason, J.B., & Fox, J.W. (1989). Amino acid sequence of a *Crotalus atrox* venom metalloproteinase which cleaves type IV collagen and gelatin. *J. Biol. Chem. 264*, 11575-11583.

Takeya, H., Arakawa, M., Miyata, T., Iwanaga, S., & Omori-Satoh, T. (1989). Primary structure of $H_2$-proteinase, a non-hemorrhagic metalloproteinase, isolated from the venom of the habu snake, *Trimeresurus flavoviridis. J. Biochem. 106*, 151-157.

Tu, A.T., Toshiaki, N., & Baker, J.O. (1981). Proteolytic specificity of hemorrhagic toxin a isolated from western diamondback (*Crotalus atrox*) venom. *Biochemistry 20*, 7004-7009.

Vasantha, N., Thompson, L.D., Rhodes, C., Banner, C., Nagle, J., & Filpula, D. (1984). Genes for alkaline protease and neutral protease from *Bacillus amyloliquefaciens* contain large open reading frames between the regions coding for signal sequence and mature protein. *J. Bacteriol. 159*, 811-819.

Willis, T.W. & Tu, A.T. (1988). Purification and biochemical characterization of atroxase, a non-hemorrhagic fibrinolytic protease from western diamondback rattlesnake venom. *Biochemistry 27*, 4769-4777.

Willis, T.W., Tu, A.T., & Miller, C.W. (1989). Thrombolysis with a snake venom protease in a rat model of venous thrombosis. *Thromb. Res. 53*, 19-29.